# SGI Altix Ultraviolet

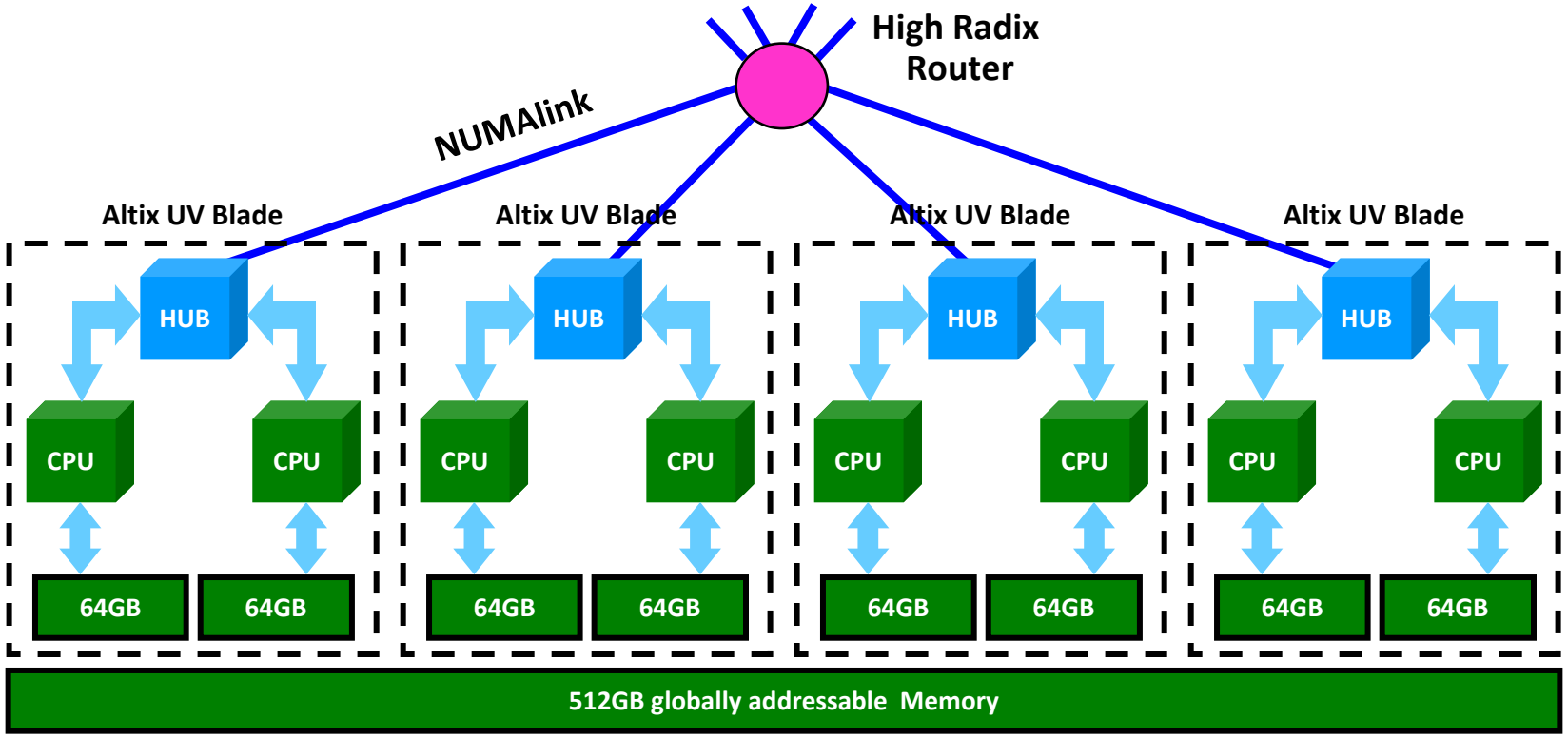
*Ruediger Wolff*
*rgw@sgi.com*

# SGI Altix UV1000

- Highly configurable system
  - SMP SSI with “large memory” datasharing
    - up to 256 Intel Nehalem-EX/Westmere-EX Sockets
    - Up to 16TB main memory
  - MPP w. up to 256 hardware connected partitions
    - Up to 32768 sockets, 16384 blades
    - Proven cross-partition communication
  - Hybrid system design
    - Two internal/external accelerator-cards per node
    - Supported accelerators
      - Nvidia GPU and GPGPU
      - AMD GPU and GPGPU
      - Tilera multicore
      - FPGAs Xilinx, Altera etc.
- High performance NL5 Interconnect
  - communications offload capability
  - 16-port routers

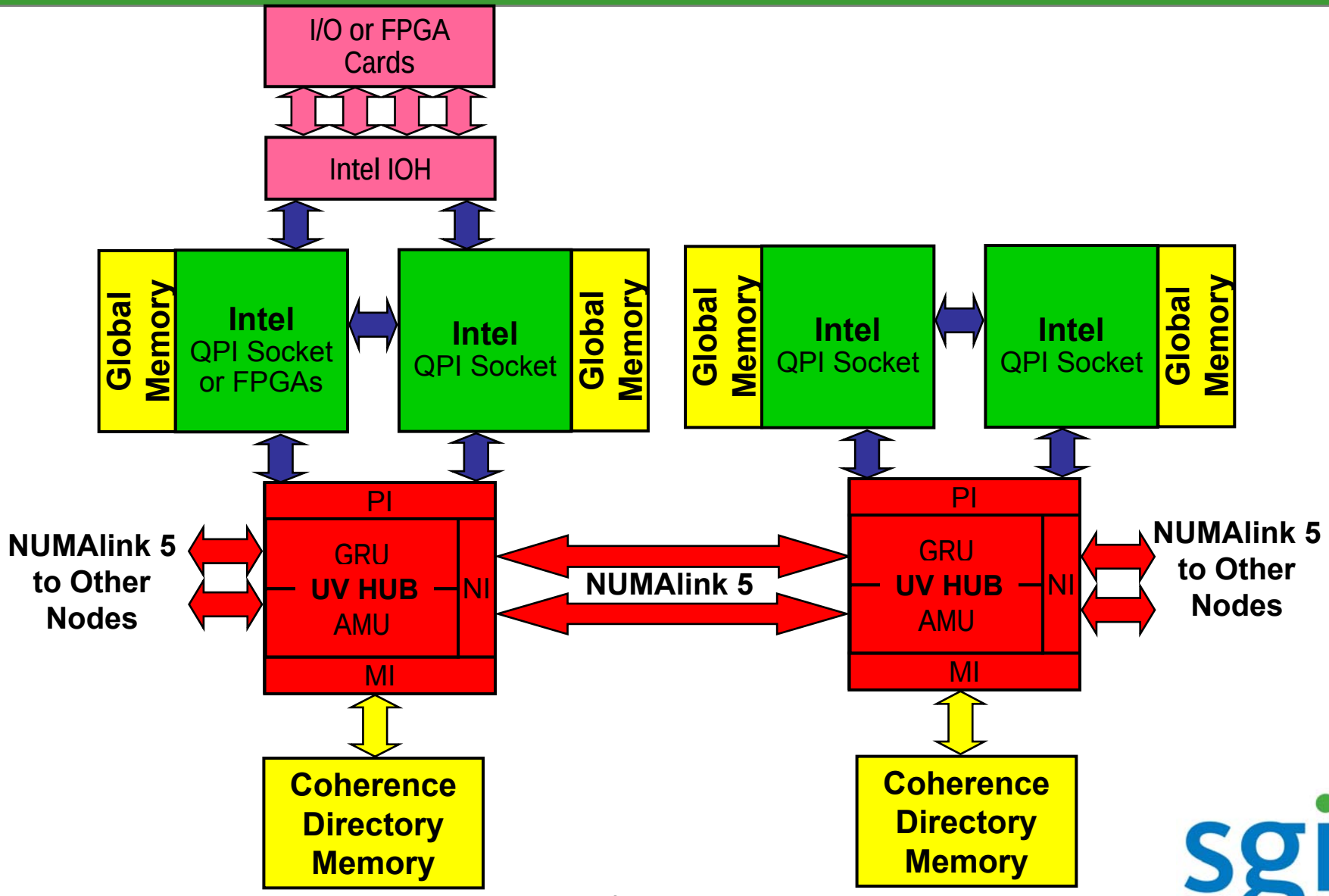
# SGI UV Interconnect with Global Addressing

NUMAlink® routers connect nodes to Multi-rack UV systems
HUB snoops Socket QPI and accelerates remote access
HUB Offloads Programming models MPI, UPC, (CoArray not yet)

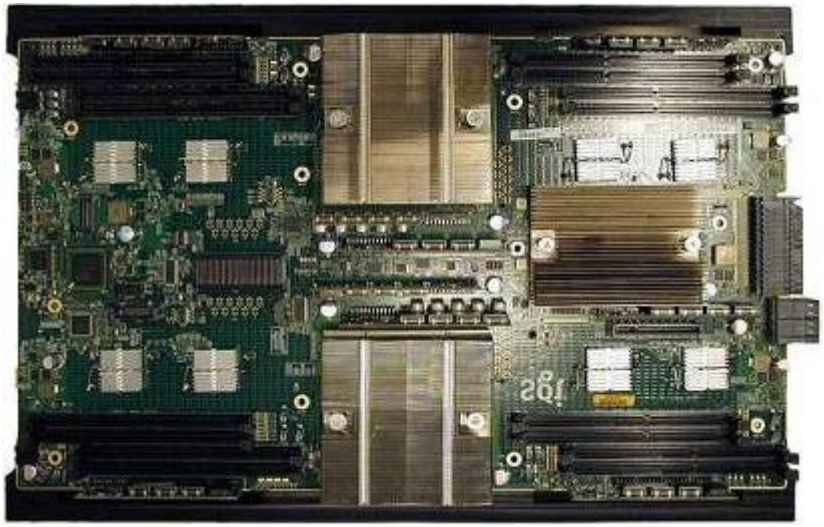


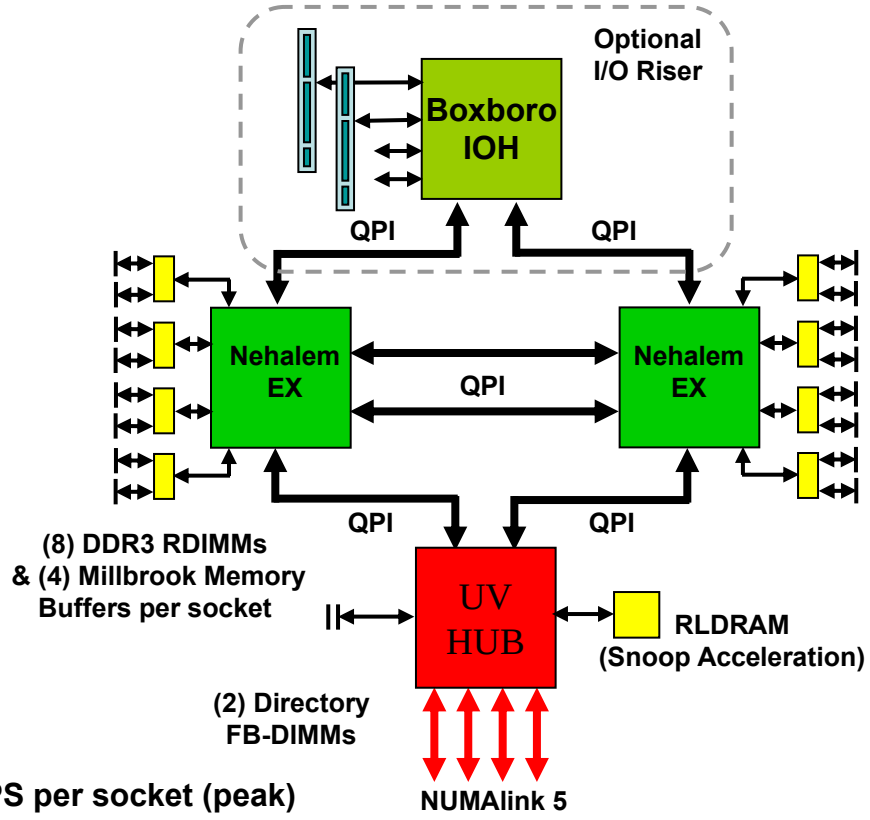

**SGI NUMA: Directory based cache coherence….table nodes that hold copies of local data**

# UV Interconnect Architecture
# UV HUB Based



# Altix UV Compute Blade





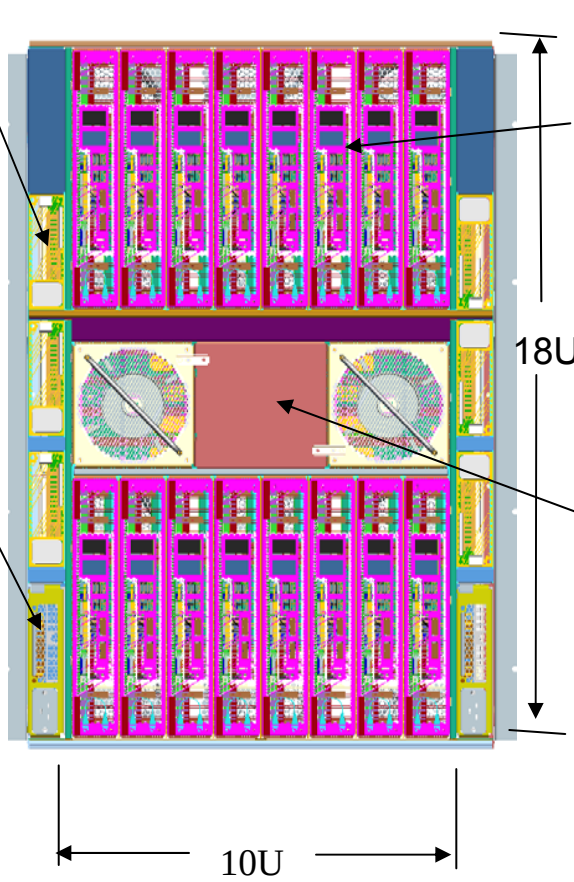

- **Nehalem-EX**
  - **8 cores per socket**
  - **2.27 GHz x 4 GFLOPs per clock x 8 cores = 72.64 GFLOPS per socket (peak)**
  - **145.28 GFLOPS per blade**
- **NUMAlink 5**
  - **15 GB/s x 4 = 60 GB/s per blade**
- **QPI = 25.6 GB/s x 4 = 102.4 GB/s per socket**
- **Millbrook Memory Buffers**
  - **8.53GB/s (1067MHz DDR3 DIMMs) x 4 channels = 34.1 GB/s Read / Socket**
- **Intel® Scalable Memory Interconnect (SMI) = 30 GB/s/socket**

# Numalink 5 Interconnect

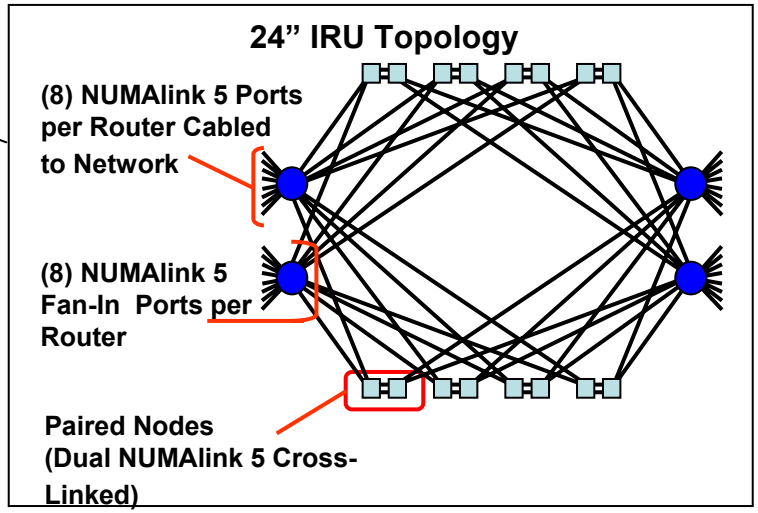
- Numalink Cables
  - Based on 12x IB-Cables
  - Linkspeed 7.5GB/s per direction, 15GB/s bidirection
  - Numaprotocol
- UV Router
  - 16 Numalink5 ports
  - Packed in groups of 4 routers
    - QCR-Quad-compact Routers
- UV Hub-chip
  - 4 Numalink5 ports
  - 2 QPI ports
  - AMU and GRU

# Blade Chassis – Independent Rack Unit

N+1 (5+1) 2.8KW PS



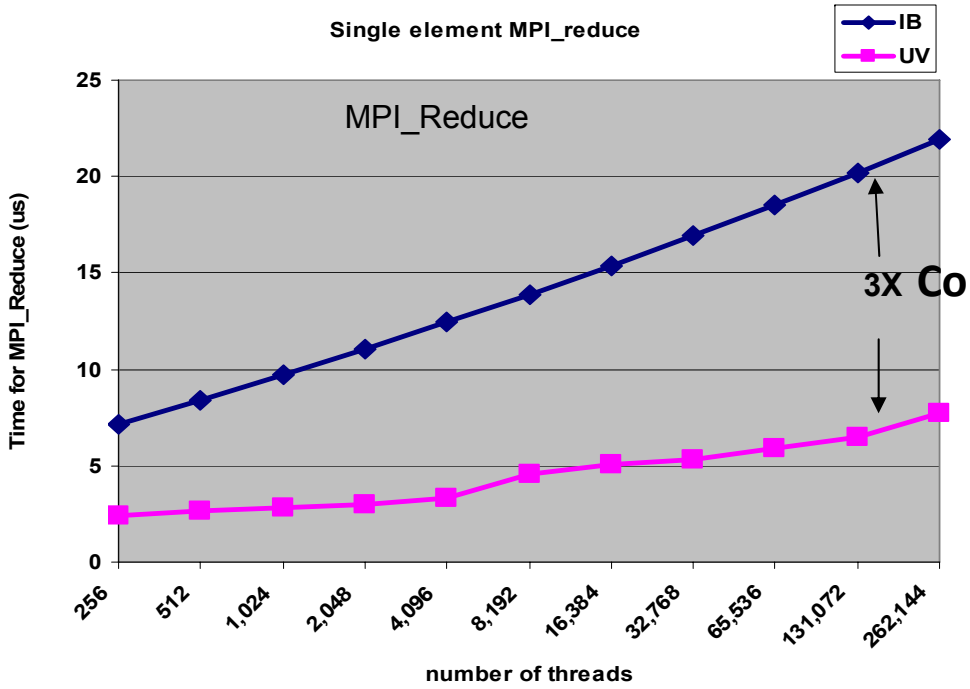

16 blade IRU for 24” rack



# UV NL5 Barrier and MPI Reduction Performance

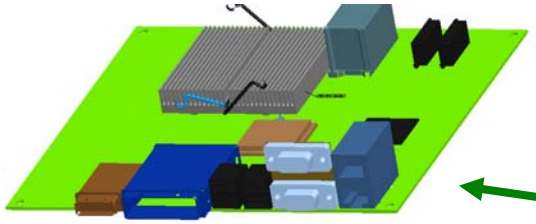
## Very high message rates, 500k barriers/s and 3x MPI_reduce over IB

| UV Barrier Latency Performance | Comparison Systems |
|---|---|
| Latency = 0.3 - 1.9us (16->4096 thread)<br>Barriers/s = 490K - 380K " " | 50 us on 64 thread (Scali)<br>7.5 us on 32 thread (Altix4700) |




**3X Collective performance over IB**

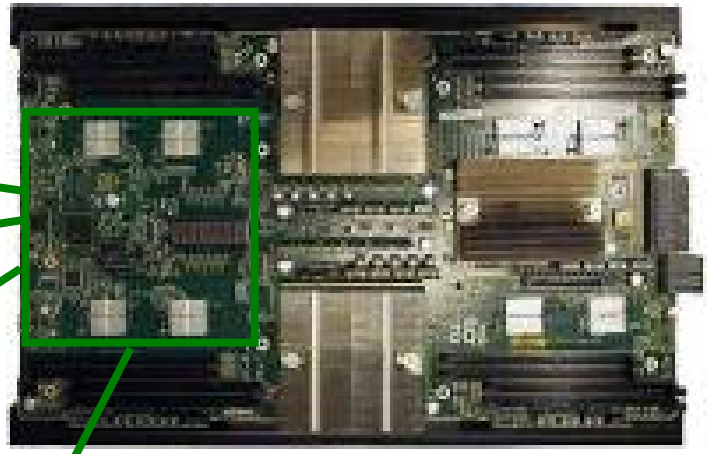
# Integrated I/O Riser Options

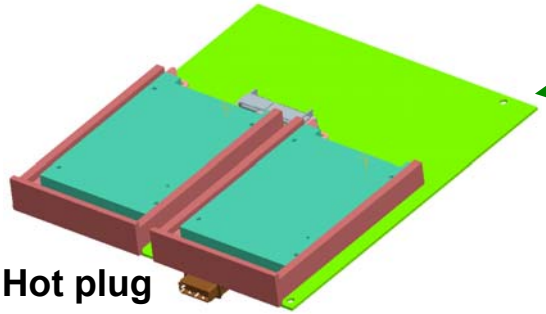


3 USB 2.0 ports
VGA controller, with 1 port
SAS controller with 2 mini-SAS 8 ports
2 GbE ports
Serial port

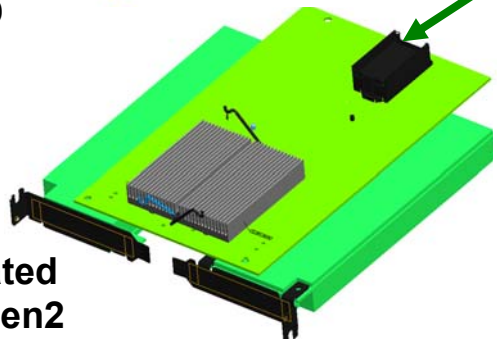
Base IO

Altix UV Blade



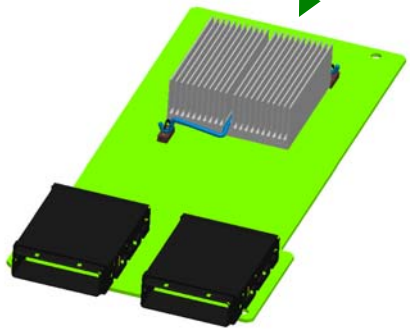


(2) Hot plug
2.5” Boot HDD



Integrated
PCIe Gen2

(1) x16 low-profile
(1) x8 low-profile



Externalized IO

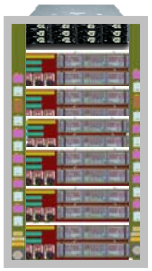
(2) PCIe Gen2 x16 Cable Connections to IO Expansion Chassis

# SGI Altix® Ultraviolet Productfamily
## Scalability with flexibility in memory and core density

- **Altix UV 1000**
  - 42U fully integrated cabinet-level solution
  - Scales to 256 sockets & 16TB shared memory (4 racks)
  - Scales to  32k sockets & 2PB global addressable memory
  - Numalink5 interconnect
  - **Up to 18.6 Tflops in shared memory**
  - **Up to 1 Pflop in global addressable memory**

UV 1000

- **Altix UV 100**
  - 3U 19” rackmount
  - Scales to 96 sockets & 6TB shared memory  (2 Racks)
  - Numalink5 interconnect
  - **Up to 7.0 Tflops in 2 racks**

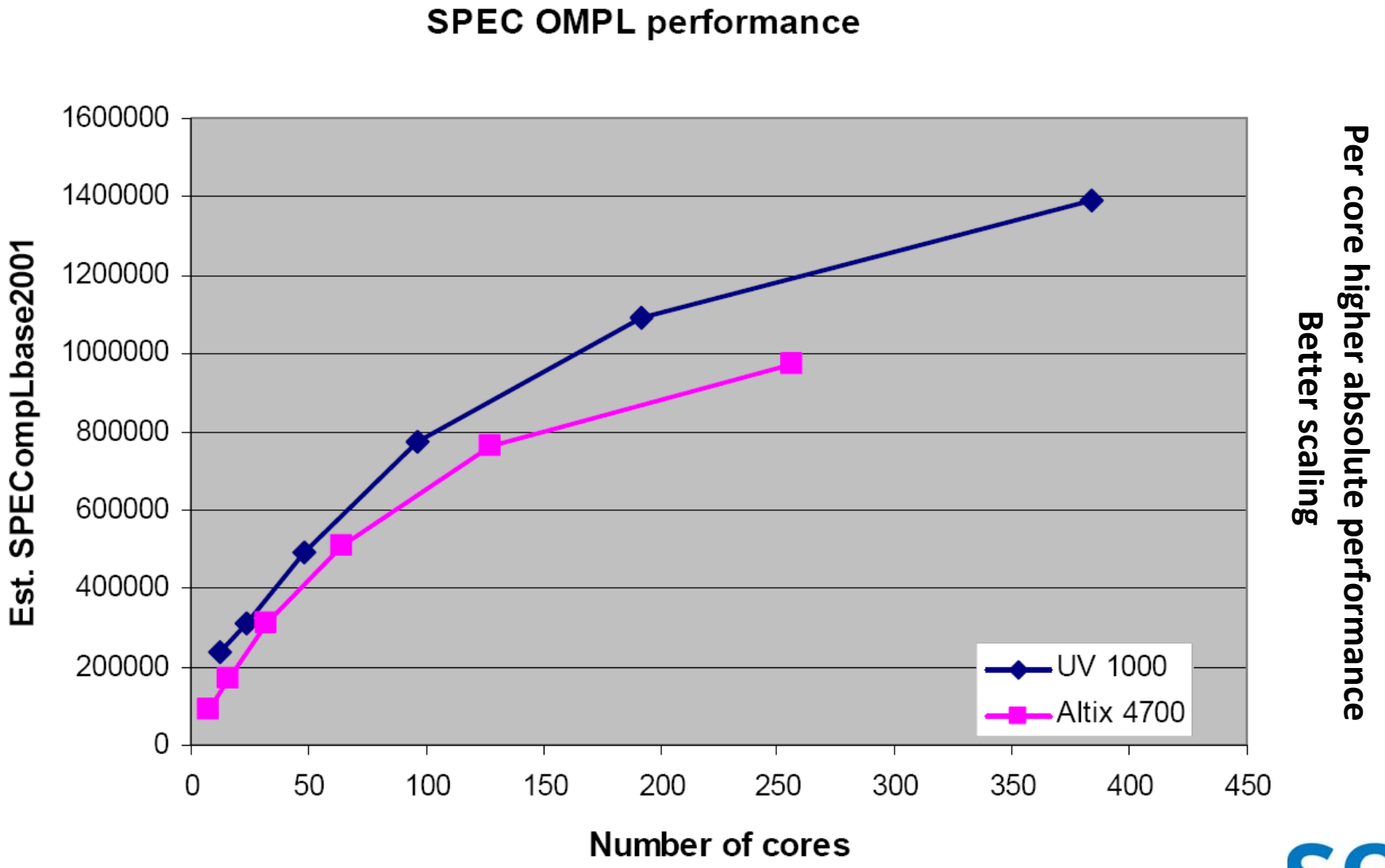


UV 100

- **Altix UV 10**
  - 4U 19” rackmount
  - Scales to 4 sockets, 32 cores & 512 GB shared memory
  - **Up to 290 GF/s**

**UV10**

# SPEC OMPL Performance



# Performance

- SPEC JBB2005

| Altix UV 1000 | | | | | |
|---|---|---|---|---|---|
| blades | JVms | threads | sockets | bops | bops/jvm |
| 4 | 1 | 48 | 8 | 1205884 | 1205884 |
| 4 | 8 | 64 | 8 | 3421167 | 427646 |
| 16 | 32 | 256 | 32 | 11919402 | 372481 |
| 16 | 128 | 256 | 32 | 12665917 | 98952 |
| 32 | 64 | 512 | 64 | 26328087 | 411376 |
| | | | | | |
| Altix4700 | | | | | |
| 64 | 64 | 256 | 128 | 5180451 | 80945 |

**Faktor 5**

**Database Search Engine Benchmark:**

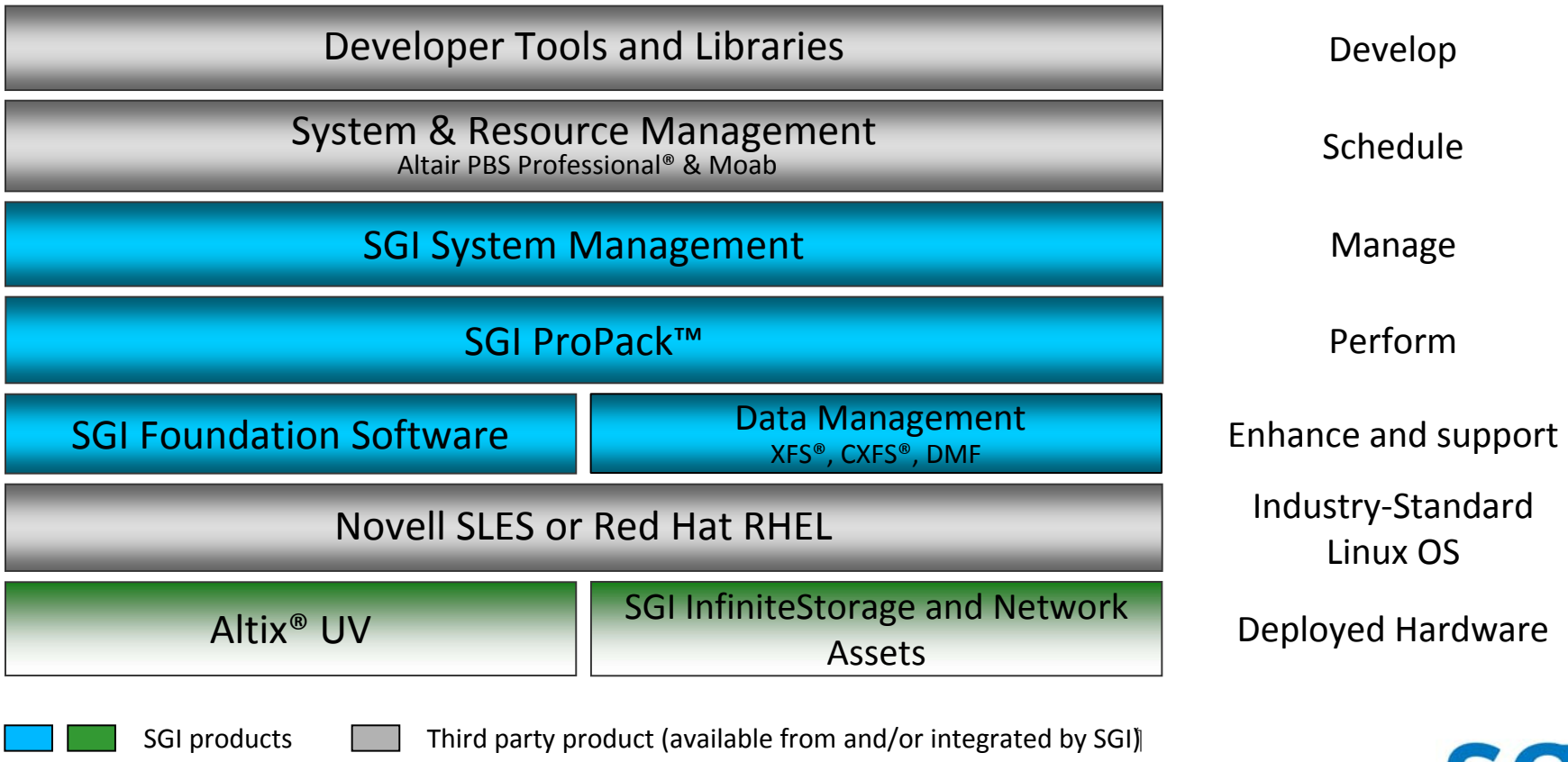
**expected improvement at the least 2x improvement**

# UV Software

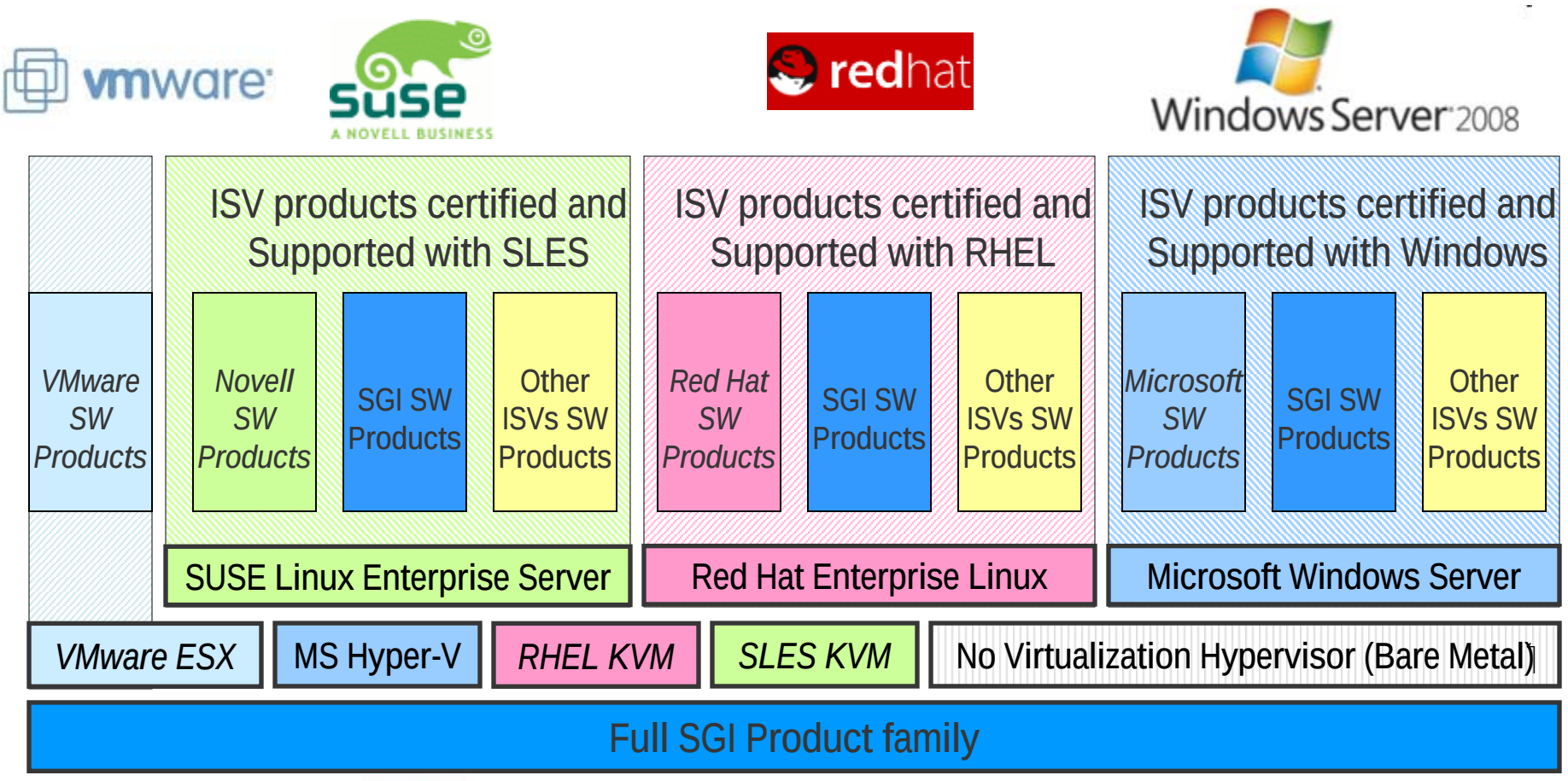
- Linux OS Community Features to Support UV
  - Key items already submitted to assure adoption by UV launch
  - SLES11 SP1, RHEL6
- Drivers, APIs
  - UV HUB/Node Controller Feature Enablement
- System Management, Integration
  - Console
  - Monitoring, debug
  - Partitioning
  - Integration with storage, data sharing across UV and other systems
- RAS – enable resiliency features of UV HUB + advanced memory RAS
- Unified Parallel C source-to-source translator
  - On Intel or GCC compiler
- Ongoing system management, MPT and other Propack advances
  - Message Passing Toolkit
  - Support for SHMEM, OpenMP

# Open: Comprehensive Software Stack

- Tightly integrated SGI software eases use and management
- X86 architecture provides access to huge array of software options



# Full Range of Fully Supported Software Environments with SGI Systems



# German UV Customers

| Institution | System | Cores | CPU Typ | GB memory |
|---|---|---|---|---|
| Konrad Zuse Zentrum fuer Informationstechnik | UV1000 | 2432 | cores(X7560) | 9728 |
| RRZN Universitaet Hannover | UV1000 | 2432 | cores(X7560) | 9728 |
| Leibniz-Institut f. Atmosphaerenphysik e.V. | UV1000 | 600 | cores(X7542) | 3200 |
| CTBTO Preparatory Commission | UV100 | 432 | cores(X7542) | 4608 |
| Leibniz Rechenzentrum | UV1000 | 256 | cores(X7550) | 512 |
| Merck KGaA | UV100 | 192 | cores(X7542) | 512 |
| Universitaet der Bundeswehr Muenchen | UV1000 | 168 | cores(X7542) | 896 |
| CvO Universitaet Oldenburg | UV100 | 120 | cores(X7542) | 640 |
| Fraunhofer Einrichtung ENAS | UV1000 | 120 | cores(X7542) | 640 |
| Heinrich-Heine-Universtitaet Duesseldorf | UV1000 | 96 | cores(X7560) | 768 |
| Access e.V | UV100 | 72 | cores(X7542) | 192 |
| DESY | UV100 | 72 | cores(X7542) | 384 |
| Humboldt-Universitaet zu Berlin | UV1000 | 48 | cores(X7542) | 256 |
| SL-Rasch GmbH | UV1000 | 48 | cores(X7542) | 256 |
| Saarlaendischer Rundfunk | UV10 | 24 | cores(X7542) | 128 |
| Leibniz-Institut f. Atmosphaerenphysik e.V. | UV10 | 24 | cores(X7542) | 128 |

**Recently won:**

**Uni. Zuerich 256 cores**

**CSCS 256 cores**

**Uni. Linz 2048 cores**

**….**


sgi
®